DIACLEU

An introduction to dialect classifications in Europe

Gotzon AURREKOETXEA,1 Ariane ENSUNZA,2 Jožica ŠKOFIC2 & Hans VAN DE VELDE3 *

University of the Basque Country (UPV/EHU)1 / Fran Ramovš Institute of the Slovenian Language (ZRC SAZU)2 / Fryske Akademy3 / Utrecht University
gotzonauurre@gmail.com / ariane.ensunza@ehu.eus / jozica.skofic@zrc-sazu.si / hvandevalde@fryske-akademy.nl

1. Introduction

Europe is the cradle of dialectology and in most European countries the study of geographical language variation has an almost century long tradition. In fact, regional variation and the dialects of all the national and most of the indigenous minority languages spoken in Europe have been the topic of scientific research. Classifications of dialects and dialect groups have been produced for each of these languages. Despite the numerous language-centred dialect studies and studies focussing on dialect variation across language borders and on larger geographical areas, such as Europe (e.g., Atlas Linguarum Europae, European Dialect Syntax), there is no publication that attempts to collect and analyse the dialect classifications in the endogenous languages spoken in Europe. DIACLEU, Dialect Classifications of Languages in Europe, proposes to fill this gap, in a series of issues that will be published in this journal.

* UPV/EHU Escuela de Ingeniería de Bilbao (Edificio I) Plaza Ingeniero Torres Quevedo, 1; 48013, Bilbao (Bizkaia).
Fran Ramovš Institute of the Slovenian Language, Novi trg 2, 1000 Ljubljana, Slovenia.
Fryske Akademy, Doelestraat 8, 8911 DX Leeuwarden, The Netherlands. Van de Velde’s work is partly funded by the COLING project, HORIZON 2020, MSCA-RISE, Grant Agreement ID 778384, https://coling.al.uw.edu.pl
© Author(s)
The roots of the project are a panel on “Dialect classifications in Europe revisited” organised by Gotzon Aurrekoetxea and Ariane Ensunza at the 10th International Conference on Language Variation in Europe (ICLave|10), hosted by the Fryske Akademy in Leeuwarden (The Netherlands) from 26 to 28 June 2019. A dozen contributions were discussed, and it was clear that the topic deserved to be treated in greater depth. Aurrekoetxea & Ensunza (UPV/EHU, University of the Basque Country, Spain), took the initiative to further develop the project, with the support of Hans Van de Velde (Fryske Akademy / Utrecht University, The Netherlands). Once the proposal became more concrete, Jožica Škofic (ZRC SAZU, The Research Centre of the Slovenian Academy of Sciences and Arts, Slovenia) joined the team.

The aim of the project is introduced in Section 2. The criteria, principles and requirements for inclusion of a language in DIACLEU are elaborated in Section 3. The characteristics of this collection of papers that will be published in *Dialectologia*, the standard structure of the papers, and the project’s accompanying website (www.diacl.eu) are described in Section 4. The rationale for adding a couple of theoretical and methodological papers is explained in Section 5. Section 6 presents an overview of the papers in this first issue.

**2. Aim**

The aim of DIACLEU is to publish a comprehensive, state-of-the-art and historical overview of dialect classifications of the indigenous languages spoken in Europe. The project collects all the classifications published from the 19th century until now for each of the languages and makes them, in English, accessible for an international audience, following a fixed format (Section 5). The structure of DIACLEU and the organization of the classifications per language facilitates comparison between (groups of) languages.

Each individual paper provides a full overview of the dialect classifications of a particular language, only taking into account published studies. The papers do not provide results of unpublished original research work and work in progress, but only
refer to these forthcoming studies. So, each contribution is a state-of-the-art reference work for that language. The format allows a continuous update of the individual papers, and updated papers can be published on the project’s website (diacl.eu). The title of each contribution follows a fixed format: language name (in English) dialect classifications. To facilitate quick reference to the languages covered, the contributions in each issue will be arranged in alphabetical order by their English language name.

In order to compare and analyse the classifications within and between languages, emphasis has been placed on the theoretical frameworks in which they have been developed. Each contribution focuses on the theoretical framework, the linguistic features used in the classifications, the methodology and – if applicable – the statistical techniques. Each classification is critically evaluated by the author(s). It is not the primary aim of this series to discuss linguistic characteristics of the dialects or dialect groups in detail. They are sometimes listed and discussed in the original classifications, but often there are no systematic descriptions of all dialects and subdialects. In this series of papers, the main linguistic characteristics used in the classifications are mentioned, to give insight in the (changing) principles and methods used in dialect classifications in Europe.

DIACLEU has a diverse target audience. Researchers in dialectology and geolinguistic variation are the primary intended users of this volume. In addition, it will be of interest to researchers in language typology, to specialists of different grammatical components (phonetics, phonology, morphology, lexicon, syntax) interested in their role in dialect classifications, and to scholars working on geographical variation in other scientific disciplines. It will also be an essential reference work for course developers and teachers of introductory classes in linguistics, dialectology, sociolinguistics, language diversity, language variation and change, language policy and language management and for courses on each of the languages. Its use is not limited to Europe, as several of the European languages are major national languages and/or popular foreign languages outside Europe. Furthermore, the inclusion of many European (lesser-studied) regional minority
languages will attract a wider scope of users. Finally, given the large number of maps, DIACLEU might be of interest for language enthusiasts.

We adhere to full and immediate open access of the papers published in the DIACLEU series. Science should be unencumbered, both for researchers and for the world-wide scientific community. Therefore, there are no publication fees for the authors and the access to the papers is free and unrestricted. At the same time, we strive for international visibility, in particular in the field of dialectology. *Dialectologia* is an online international journal that meets our criteria, and we are happy that they accepted to host the publication of DIACLEU in a series of special issues under a CC BY-NC-ND licence.

3. Criteria, principles and requirements for inclusion in DIACLEU

DIACLEU a priori intends to include all indigenous languages currently spoken in Europe. It covers different language families or groups such as Abkhazo-Adygian, Baltic, Celtic, Eskaleut, Finnic, Germanic, Iranian, Kartvelian, Nakho-Dagestanian, Permic, Romance, Semitic, Slavic and Ugric, as well as contemporary language isolates such as Albanian, Basque and Greek. It considers both national and regional/minority languages. As the focus is on spoken languages, sign languages – which have furthermore not been studied from a dialectological perspective – are outside the scope of this project. Special attention has been paid to the issue of languages spoken across national borders, most of them sharing a dialectological tradition dating back to the 19th century. However, some of the present-day European languages (e.g., Aragonese in Spain) have been considered as varieties of another language until quite recently, and in several cases (e.g., Macedonian, Croatian), this is still a sensitive and controversial political and linguistic issue the editors and contributors have to deal with. These issues will be addressed in more detail in the following issues.

In Section 3.1 we present the criteria to be possibly included in this series. The list of languages is further reduced by means of requirements (3.2). In Section 3.3 we
describe two principles used in the composition of this series of papers on dialect classifications.

3.1 Criteria

Considering the aims of DIACLEU, three basic criteria have been established: (i) to be an indigenous European language, (ii) to be part of the Atlas Linguarum Europae – ALE project, and (iii) to be in the list of regional / minority languages of The European Charter for Regional / Minority languages. All languages in this volume must fulfil the first criterium. The second and third criterium are used to complete the list of candidate languages.

3.1.1 To be an indigenous language spoken in Europe

In a geographical sense, Europe is defined as the continent bordered by the bodies of water in the north (Arctic Ocean), west (Atlantic Ocean) and south (Mediterranean Sea). In the east it borders with Asia. This border is formed by the Ural Mountains and Ural river, the Caspian Sea, the Caucasus mountains, the Black Sea and its waterway connection with the Aegean Sea. Traditionally, geographers group islands with the nearest continental landmass (e.g. the Azores, Faroe Islands, Great-Britain, Iceland, Ireland, Malta, Sicily). We expand this geographical entity on the basis of political arguments and define Europe as a political-geographical entity. Greenland is geographically assigned to North-America, but given its geographical bordering to what is traditionally seen as Europe, its political belonging to Denmark, and the recognition of Greenlandic as one of the languages of the Danish Kingdom, it is considered as a part of Europe in this project. For Turkish we decided to focus only on the dialects spoken in the European part, including Turkish spoken on Cyprus, which is politically and culturally considered as part of Europe. The following nation states are
part of Europe: Abkhazia, Albania, Andorra, Armenia, Artsakh, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kazakhstan, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Republic of Moldova, Monaco, Montenegro, Netherlands, North Macedonia, Northern Cyprus, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Slovak Republic, Slovenia, South Ossetia, Spain, Sweden, Switzerland, Transnistria, Turkey, Ukraine, United Kingdom, Vatican City.

We focus on the languages that have been indigenous to these European countries for many centuries. So, heritage languages (mainly introduced to immigration in the 20th and 21st centuries) and European languages spoken outside Europe (mainly as a result of colonization) are excluded from this volume. The territories that are still politically attached to a European country (e.g. the French overseas territories) and that are not part of the traditional studies in spatial dialectological are outside the scope of this volume. According to the Council of Europe, Europe has about 225 indigenous languages, and almost half of them are spoken in Russia. However, they do not provide a full list of languages. On our website we provide a list of the indigenous languages in Europe that are the basis of this project and how these languages fit the other criteria and principles and for which languages authors are working on papers.

---

1 Entities with only partial diplomatic recognition. None are members of the Council of Europe, the Europe Union or the United Nations. These entities are colored in grey in Map 1, which does not mention their name.
2 Not a member of the Council of Europe.
3 Russia (the Russian Federation) was suspended as a member of the parliamentary assembly of the Council of Europe on 15 March 2022. The same day Russia withdrew from the Council of Europe.
4 Vatican City or Holy See is not a member of the Council of Europe, but is an observer state.
3.1.2 Inclusion in the *Atlas Linguarum Europae*

The *Atlas Linguarum Europae* (ALE) was founded in 1970 (Weijnen, Hagen & Kruijzen 1972) and the frontiers of this linguistic atlas are neither political nor linguistic, but geographical (https://lingv.ro/atlas-linguarum-europae/). This established European project has sufficient scope and prestige to be the starting point for the selection of candidate languages to be included in this new European project in the field of dialectology, as the geographical scope of ALE almost completely overlaps with our scope, defined in Section 3.1.1 and the focus is also on (documented) dialects and spoken language varieties. As far as we know, only Armenia, Azerbaijan (and Artsakh), Greenland and Northern Cyprus\(^5\) were outside the geographical scope of ALE.

\(^5\) ALE only lists Greek dialects on Cyprus, not the Turkish ones spoken in Northern Cyprus.
In its most recently published overview, ALE presents an overview of about 100 languages and dialects, without any distinction between official and minority languages (ALE 2015: XXI). However, it should be noted that some of these are a cover term including several languages, such as gallo-roman (including Franco-Provençal, French (Oïl), Occitan (Oc)) and italo-roman (including Corsican, Friulian, Italian, Ladin, Sardinian). The table on our website gives an overview of the languages that are included in ALE.

3.1.3 Recognition of the language under the European Charter for Regional or Minority Languages

We aim to contribute to the visibility and scientific study of regional and minority languages, by treating them on equal footing as the national and official languages in Europe. Therefore we add a third criterium: recognition under the European Charter for Regional or Minority Languages (ECMRL). In November 2023, only 23 of the 46 current member states of the Council of Europe have recognised regional, minority or non-territorial languages under this charter, covering a total of 81 languages. Two member states (Liechtenstein and Luxembourg) have ratified the ECRML, but declare that no regional or minority languages are spoken on their territory.

Two types can be distinguished: languages that are recognised under the charter in one nation and have a higher official recognition in another nation (e.g., Polish in Hungary vs. Poland, Italian in Slovenia vs. Italy); and languages that are a minority, regional or non-territorial language in all regions or nations where they are spoken (e.g., Basque, Catalan, Frisian, Romani, Yiddish). The first type overlaps almost completely with the ALE list, the second type adds numerous new languages.

The fact that not all European countries have ratified this charter (e.g., Belgium, France, Greece, Italy, Portugal), might still result in an incomplete list of languages and a linguistic project of this type should not be hampered or biased by political obstacles.

6 They list four Italian regional dialects: Gallo-Italian, North, Middle and Central Italian. And languages that are also spoken outside Italy: Albanese, Croatian, German, Greek and Slovene.
However, given the focus on two centuries of research in dialectology, we are confident that we didn’t miss any of the dialect classifications in the regions we are familiar with. But there is big geographic-linguistic gap in the expertise of the editorial team: a specialist of the non-Slavic languages spoken in the north-eastern part of Europe.

3.2. Requirements

In addition to the criteria mentioned above, four requirements have been established that must be met in order to be included in the project: the language in question must have at least one published dialect classification, it must have native speakers, be used in daily communication and linked to a territory (in Europe).

3.2.1 The existence of dialect classification studies

As a first requirement, the language must have been the topic of scientific dialectological research and have an established bibliography on dialect varieties. In addition, the existence of at least one dialect classification, whatever its nature, quality or theoretical framework, is a prerequisite to be included as a language in the DIACLEU project.

3.2.2 Spoken by native speakers

As a second requirement, the language must have native speakers. Languages that only have second language speakers, such as Manx on the Isle of Man and Latin in Vatican City are excluded from this project.
3.2.3 Used in daily communication

The languages included in this project are used in daily communication. Languages that are restricted to a very specific domain (e.g., religion, rituals) will not be dealt with.

3.2.4 Linked to a territory

In this project on spatial dialect variation languages that are not directly linked to a specific territory in Europe, i.e., languages of nomads (e.g., Romani) and dispersed languages (e.g., Yiddish) will not be included.

3.3 Principles

Two general principles will be followed in the DIACLEU project: (i) one paper per language and (ii) dialectological traditions. It should be noted that these principles can be conflicting. Furthermore, in some cases the official status of languages has shifted over time due to political changes. In other cases, dialectological traditions have changed, sometimes related to political changes.

3.3.1 One paper for each language

As a first principle, each language will be treated in a single overview paper of its dialect classifications, independently from its geographical spread over several nations, differences in its legal or social status across administrative borders. So, German in Austria, Belgium, Germany and Switzerland are all part of the paper on German dialect classifications. The paper on Dutch will cover the dialects of Dutch spoken in the Netherlands, Belgium, and France. The situation is similar for minority languages: Basque in Spain (recognised in two autonomous countries) and France (without any official recognition) are also treated in a single paper on Basque dialect classifications, the paper on Sámi covers all its varieties in Finland, Norway, Russia, and Sweden. It
should be noted that dialect classifications of varieties of European languages spoken outside Europe will not be discussed (e.g., English in New Zealand, Portuguese in Brazil, French in Canada, Turkish in Asia).

3.3.2 National or language-specific dialectological traditions

As a second principle, we attempt to respect the traditions of dialectology in each nation and for each language. Sometimes different – genetically related – languages have been treated together in spatial dialectology. For example, in France, the French, Occitan and Franco-Provençal languages spoken in France, Switzerland, Italy and Belgium have been studied jointly as the Gallo-Romance dialects and this volume follows this tradition of French dialectology. The Low-Saxon and Limburgian dialects spoken, now recognised as regional languages in the Netherlands under the ECRML, are still part of the Dutch dialectological tradition and hence not treated as separate languages. But traditions can change over time, as is the case for Luxembourgish, where dialectology started as part of the larger spectrum of German dialect studies, but since the 1960’s has changed towards a dialectological research tradition considering Luxembourgish as a separate language (see Gilles, this issue). For some of the Slavic languages similar tendencies are observed. With this approach we intended to offer the reader a better understanding of the dialect classifications of these languages.

4. Characteristics of this collection of papers

All indigenous languages in Europe that meet the above criteria and requirements have a place in this project. Following our principles, each language is treated equally, regardless of the size of the language area, the number of speakers or the official status of the language. The project is governed solely by the above-mentioned characteristics. Contributions for about fifty languages are in progress, but
we invite scholars from all over Europe, and especially from the languages spoken in the north-eastern part of Europe to contact us and propose a contribution.

DIACLEU offers two types of publications: (i) static papers in *Dialectologia*, an online open access journal (www.edicions.ub.edu/revistes/dialectologia/), that will be published in four special issues in 2023 and 2024, and (ii) a dynamic website (diacl.eu).

The website started as an information portal providing information for prospective authors, such as information about the aim of the project, a style sheet and a list of scientific terms. The full list of languages and authors can also be found on the project’s websites. For some languages we are still looking for authors. If a language is not mentioned in the list, but you think that it meets the criteria, principles and requirements, we invite you to contact us. Once the papers are published in *Dialectologia*, they will also become available on diacl.eu. Furthermore, updated versions of the papers will be published (in the case of new classifications) and information in relation to dialect classifications in Europe will be presented. The purpose is to create a network of international scholars working on dialect classifications, also outside Europe, and to support and stimulate future work on this topic to give greater visibility to studies of geolinguistic variation in Europe.

All languages are treated equally in this volume, independent from the size of the language area, the number of speakers or its status. To make the classifications more accessible and comparable for the readers, all papers follow the same structure. In the introduction (Section 1) a brief historical and present-day overview of the socio-political and sociolinguistic context of the language is sketched, including the language’s geographical spread (supported by a map), its official status and number of speakers. The introduction presents for each language the basic information that is needed to understand the context of the classifications, which vary a lot across languages and areas. In Section 2 the dialect classifications are presented in a chronological order. This enables the reader to see more clearly the historical development of the dialect classifications in each language separately, how they build upon each other or how they try to provide alternative insights, often inspired by changing linguistic insights, new research methods or different or changing language policy aims. There is a subsection for each classification, consisting of two parts: (i) the
theoretical framework and (linguistic) features used in the classification and (ii) the classification in dialects and subdialects. The names of the languages, varieties and dialects are given in the language itself and/or the language of the original publication, and – if possible – in English. Names of varieties in non-roman script are also transliterated to roman script. The chronologically presented classifications are followed by a discussion in which the authors of the papers focus on the strengths and weaknesses of the classifications, their contribution to language policy and linguistic debates in the language area, and to dialectology in general. Most authors also focus on the future of the field in their language area. At the end of each paper a reference list is provided.

4.1 Theoretical framework

In this project classification of approximately fifty languages will be presented. These languages from several language families do not only differ in number of speakers and geographical space, but they also belong to different dialectological traditions and for many of these languages a time period of two centuries is covered. Furthermore, dialect classifications have sometimes served different language political purposes. Hence, a large number of theoretical frameworks have been used in the classifications of the dialects of Europe. To enable a systematic comparison within and between languages in this collection of papers, the classifications are grouped in four main frameworks, based on (i) extralinguistic criteria (ethnological classification), (ii) a qualitative linguistic analysis of language data (isoglottic classification), (iii) layman’s perceptions of geographical differences in speech (perceptual classification) and (iv) a quantitative analysis of linguistic data (dialectometric classification). These four classification types will be used as a structuring principle throughout the collection of papers, but more specific information on linguistic theoretical frameworks (e.g. neogrammarian, structuralist, generative, sociolinguistic) will be given in the respective sections.
4.4.1 Ethnological classification

The category of ethnological classifications includes classifications that are mainly based on non-linguistic features, such as cultural differences, ethnic background of the population, settlement history, geographical regions, political boundaries or on lay man's observations. These classifications were mainly made by non-linguists, such as anthropologist, ethnologists and historians (Caro Baroja 1945 for Basque, Merlo 1924 for Italian, etc.), before the rise of the field of dialectology and the collection of linguistic atlases.

4.1.2 Isoglottic dialectology

With isoglottic dialectology we cover all the classifications based on qualitative dialectological studies, both diachronic (historical, genealogical, comparative linguistics...), and synchronic (structural, functional, generative, sociolinguistic, ...), referring to all classifications based on qualitative analyses of spatial dialect data (often collected for linguistic atlases), without using quantitative research methods. The geolinguistic point of view is at the heart of the classifications and the geolinguistic approach in dialectology (in contrast with monographic studies) has been labelled as linguistic geography, dialect geography, spatial dialectology, geolinguistics or regional dialectology. The term feature-based classification refers to the fact that the classification is based on linguistic features, commonly defined by experts (dialectologists, linguists). However, we opt to introduce the term isoglottic dialectology as the isogloss concept (phonologic, lexical, morphologic, syntactic or semantic) is the unifying factor of all these approaches.

4.1.3 Perceptual dialectology

Perceptual dialectology is the study of nonlinguists’ views of spatial variation within languages. A wide range of techniques have been used to collect these views and both qualitative and quantitative methods have been used to analyze them.
4.1.4 Dialectometry

Dialectometry uses the input from traditional spatial dialectology (especially linguistic atlases), but instead of a qualitative classification, quantitative, mathematical, and statistical techniques are used to analyze large data sets.

4.2 Classification of dialects and subdialects

For each classification information is provided about the classification method and – if available – on which grounds dialects and subdialects were defined, how dialect boundaries were established and whether transition zones were distinguished by the original author(s). In addition, insight will be given in the main features used in the classifications (e.g., extralinguistic, sociolinguistic, lexical, phonological, morphological, syntactic, prosodic), Within a given language, the different classifications will be evaluated, and their development will be discussed in the light of their aims, sociolinguistic context, and developments in linguistics and dialectology.

5. Theoretic-methodological contributions

This overview of dialect classifications in European languages contributes to the knowledge about the diversity of languages, linguistic situations and the position of languages and dialects, enriching the panorama of European linguistic diversity. It tries to assemble these classifications in different theoretical frameworks and make them accessible for an interlinguistic comparison about the research methods, the resources used, the characteristics taken into account in the classification, the regional variation in languages, etc.

We also try to look forward. We invited European scholars to reflect on methodological and theoretical issues in geolinguistics. We hope these papers will contribute to the strengthening of the theoretical basis of dialect classifications and to
an improvement of the current classification methods. These contributions will hopefully open new paths and be a source of inspiration for new projects. In this way, we hope that the classifications, which offer primarily insights in the history of dialectology and geolinguistics, will offer a glance in the future of the study of spatial variation in languages.

6. Short introduction to the first issue

This is the first special issue of *Dialectologia* devoted to the dialect classifications of languages in Europe. Following this introduction, this issue contains one methodological paper on linguistic atlases and dialectometry (Hans Goebl) and nine papers on different languages. In this issue there are papers on Basque (Ariane Ensunza & Lorea Unamuno), Finnish (Marjatta Palander), Gallo-Roman (Guylaine Brun-Trigaud), Greenlandic (Michael Fortescue), Irish (Raymond Hickey), Italian (Federica Cugno), Luxembourgish (Peter Gilles), Norwegian (Agnete Nesse) and Welsh (Gwenllian Awbery). The papers are presented in alphabetical order by English language name.

The first issue already illustrates the geographical space and the linguistic and sociolinguistic diversity covered by this project. The languages are members of the Celtic (Irish, Welsh), Eskaleut (Greenlandic), Germanic (Luxembourgish, Norwegian), Romance (Italian, Gallo-Roman languages) and Uralic (Finnish) language families or groups, and a language isolate (Basque). The languages also differ in official status. Most of them are official or national languages in one or more nations: Finnish in Finland, French (as part of the Gallo-Roman paper) in Belgium, France, Luxembourg and Switzerland, Irish in the Republic of Ireland, Italian in Italy, Luxembourgish in Luxembourg, Norwegian in Norway. Greenlandic has official status in Greenland, an autonomous part of the Danish Kingdom. Basque is spoken in parts of Spain and France, but only recognized as a minority language in parts of Spain (Basque Country and Navarre). Welsh is recognized as a minority language by the United Kingdom.
National of official languages can also be recognized as minority languages in other countries, e.g., Finnish in Sweden, Irish in the United Kingdom. Italian in Bosnia and Herzegovina, Croatia, Romania, Slovenia and Switzerland. Franco-Provençal (Gallo-Roman) is recognized as minority languages in Switzerland, but not in France. Some have a large number of speakers (French, Italian), others have very few native speakers (Greenlandic, Irish), and some are endangered languages in a revitalization process (Irish, Welsh).

References

